Z - 153E
c.2.

ANALYTIC USES OF SURVEY DATA: A REVIEW

D. Binder
J. Kovar
S. Kumar
D. Paton
A. van Baaren

Statistics Canada
Ottawa, Canada
April 1985

Number: CHSM 85-082E

Abstract

A brief history of the use of survey data in data analysis is given. Issues arising from the complexities of the survey design are highlighted, both from the point of view of estimation as well as for performing hypothesis tests and developing confidence intervals. Many problems such as analysis of linear models and log-linear models have been studied recently. We cite the main results. Estimation of population percentiles is also discussed.

Keywords: Complex survey designs; Confidence intervals; Descriptive statistics; Hypothesis tests; Linear models; Log-linear models.

1. INTRODUCTION

The use of survey data using scientific sampling methods can be traced back to the late 19th century. However, until the 1960's the major emphasis of sample surveys was directed toward estimation of means, totals, proportions and ratios. With the advent of high speed computers, the number of surveys increased substantially and users became more sophisticated in their requirements. Over the last 25 years there has been a flurry of activity in the use of surveys for analytic purposes rather than purely as a descriptive device. Age-old questions such as use of sampling weights and making inferences for finite populations resurfaced in this new context. Many of these questions have now been partly resolved, at least for the case of large scale surveys where asymptotic results can be applied. In this paper we review these developments.

In Section 2 we briefly summarize the developments of survey sampling to 1970. In Section 3 we discuss the problems introduced by using surveys for analytic rather than descriptive studies. In Section 4 we discuss problems which are more particular to analysis of categorical data. Section 5 gives a brief description of problems associated with percentile estimation.

2. EARLY DEVELOPMENTS

The concept of generalizing from a part of the population to the whole was not considered on a scientific basis until the latter half of the 19th century.

In this respect there are two main issues[1]:

1. How to select the part from the whole.
2. How to generalize from the selected part to the entire population.

Whereas the focus of this paper is on the second issue, it should, however, be borne in mind that the two issues cannot be completely separated.

The first statistician who can be credited with giving sampling a more scientific basis was A.N. Kiaer, who was the Director of the Norwegian Bureau of Statistics during the last quarter of the 19th century. It is not known whether surveys taken before that time may be called sample surveys because statisticians never gave an account of how they took their samples and what controls they used. Kiaer was the first to consider it fundamental that the accuracy of a survey depends not so much on the larger or smaller number of observations made, but on the method of obtaining correct representativeness. His was the concept of the "representative method of investigations", about which he wrote several papers during the last 10 years of the 19th century. It was, however, not until 1903 that his ideas were tentatively accepted by the International Institute of Statistics. Kiaer also applied the principle of stratification, which he called the "method of controls". (It should be noted that during the same period in the United

[1] Actually there is a third issue "How to properly use the generalizations for practical purposes once they have been obtained". With a few notable exceptions this issue is ignored in many statistical textbooks.

States, C.D. Wright, who had a position in Massachusetts similar to Kiaer's, also applied the concept of representativeness which he independently developed, although on an even more intuitive basis than Kiaer did). Kiaer's ideas triggered the development of a more theoretical basis for survey sampling. L. von Bortkiewitz was the first to express the idea of testing sample results for representativeness, although he was not the first to employ statistical tests in connection with sampling. (The first known application of a statistical test had already taken place about two centuries earlier by John Arbuthnot, Queen Anne's physician.) A.L. Bowley proposed to use the theory of probability in connection with survey sampling and, building on the work of K. Pearson and F.Y. Edgeworth, developed the statistical theory of sampling.

A major milestone in the development of sampling was Neyman's (1934) paper, where he placed the methods of stratified and group sampling on a sound theoretical basis and criticized the method of purposive selection as described by Bowley and tested by Gini and Galvani.

Until Neyman's paper, the theory of estimation was mainly that of point estimation based on Bayes "Theorem of Inverse Probability" which requires the knowledge of the prior probabilities associated with different admissible hypotheses. Neyman described and refined the theory of "interval estimation" as suggested by R.A. Fisher with parameter estimates based on the Gauss-Markov Theorem. This method of estimation became the recognized theory of estimation. In the meanwhile, Fisher had also developed the analysis of variance, which is a technique to separate total variance into its various components. The validity of the error estimates is based on the principle of

randomization. It was due to the work of A.R. Clapham that the analysis of variance also proved useful for sample surveys.

After Neyman's paper, considerable developments have taken place in sampling theory, especially with respect to sample selection and estimation leading to modern sampling techniques, such as cluster sampling, systematic sampling, two-stage and multi-stage sampling, etc. All the development in this respect is covered in the five "standards" which appeared in the late forties and early fifties within five years: the books by Cochran (1953), Deming (1950), Hansen, Hurwitz and Madow (1953), Sukhatme (1953), and Yates (1949).

An important development at this stage, which had a high impact on analytic use of surveys in later years, is the use of supplementary information. Cochran studied this in connection with estimation leading to utilization of this information in the form of ratio or regression estimators. The supplementary information is comparable to the use of one or more covariates in experimental design. Before, the development was geared towards sampling variance models as introduced by Neyman (based on the Gauss-Markov theorem). Ratio and regression estimators were aimed at increasing the precision of the estimates. With the same objective in mind, Hansen and Hurwitz suggested to use additional information at the sampling stage leading to designs in which elements are selected with probabilities proportional to some suitable measure of size.

The earlier statistical methods were based on the assumption of simple random sampling and were developed from the concept of the hypothetical population model. The principle of randomization, as developed by Fisher for experimental design, was later adapted for the purpose of

survey sampling in which it implies the use of some artificial mechanism such as a random number table. This in turn implies that the elements in the target population carry some form of label, which is the case with most populations of interest in survey sampling. The statistical methods used were still based on the original assumptions of simple random samples, although in many cases they were modified to allow for sampling without replacement from a finite population. Estimates for regression coefficients and procedures such as testing for independence in a two-way table were relatively straight-forward, especially when the sampling fractions were small, since the classical procedures for analysis seemed to apply to these situations without modification.

However, in the early 1950's because of operational and efficiency concerns, more consideration was being given to stratified sampling designs and to designs with unequal inclusion probabilities; see, for example Horvitz and Thompson (1952). In the developments following, the sampling variance and its estimation still played a dominant role. The quantity of interest for this purpose, which was later called the design effect (deff) by Kish and Frankel (1974), is the ratio of the variance associated with a given survey design to that of simple random sampling.

The introduction and expansion of the use of electronic computers has contributed considerably to the development of methods of analysis of complex designs, whether survey designs, experimental designs or others, which were at one time considered to be too expensive computationally. An area of major development as a result of this introduction was regression and correlation analysis or more generally least-squares analysis. Although the method of least-squares was already proposed in the beginning of the previous

century, development of this method of analysis was slow, mainly due to the lack of means to carry out the substantial amount of computations involved, even with a moderate number of variables. It had already been recognized that difficulties may be encoutered when using regression analysis for survey data. For instance the relationship between household savings and household income using survey data was studied by Klein and Morgan (1951) who summarized the difficulties associated with: (1) weighting of observations, (2) heteroscedasticity, (3) non-linearity, (4) the choice of alternative economic concepts and (5) errors of observation; of which (1), (2) and (5) could be associated with the use of complex surveys. Konijn (1962) discussed two models which can be used for both point and variance estimation in regression relationships, where the data have been obtained using more complex survey techniques than simple random sampling, specifically stratified and cluster sampling.

For estimating the variances of non-linear statistics such as ratios, correlations and regression coefficients, one of three methods is still commonly being used: the delta-method or Taylor linearization (see Tepping; 1968), balanced repeated replication and jackknifing (see Section 3.3 for more details).

For analysis and summaries of categorical data, all the concepts introduced above could be equally applied to this special case. From the point of view of survey samplers little distinction was made between quantitative and qualitative data since the probability mechanism was applied to the population labels, not the population values. Ratio estimation of population proportions was a common way to incorporate auxiliary data beyond the design variables. Tests for proportions and for differences of two

proportions were relatively straight forward using Taylor linearization, balanced repeated replication or jackknifing methods. It was not until the early 1970's that detailed consideration was given to tests of independence, goodness of fit tests and other χ^2-type tests (see Section 4). Of course, for simple random samples, the usual χ^2 statistics were valid up to the finite population correction factor.

A major development in the analysis of categorical data from complex designs appeared in Grizzle, Starmer and Koch's (1969) paper. They generalized tests of linear and log-linear hypotheses on the estimated proportions for categorical data, based on the Wald statistics using Taylor linearization. Many of the later developments in categorical data analysis were extensions of this fundamental contribution.

3. REGRESSION AND RELATED ANALYSES

3.1 Descriptive vs. Analytic Use of Survey Data

Because of the recent concern for tests of more complicated hypotheses, such as those described by Grizzle, Starmer and Koch (1969), a distinction between "descriptive" and "analytic" use of survey data was being made. According to Kalton (1983), analytic uses refer to estimation of parameters of a causal system, ideally one of universal application. Unfortunately, in many disciplines (e.g. behavioural sciences) causal models are still in its infancy. Therefore, good predictive models, which are descriptive in nature, are sought instead. As Imrey, Sobel and Francis (1980) point out, statistical inference based on variational or structural modeling are often undertaken in several disciplines, naively assuming probability distributions based upon simple

random sampling. An example of using survey data explicitly for prediction purposes is given in Cassel, Särndal and Wretman (1979). They use the design information in their estimation, the underlying assumption being that any missing variables in a true causal model would behave similarly at a future time point.

The controversy that has arisen in survey data analysis, as it relates to regression and other linear models, extend to other analytic uses of survey data as well. The remainder of this section will concentrate on the regression framework.

3.2 Use of Sampling Design Information

For designs which are not self-weighting, the usual first question that faces the analyst who wishes to estimate certain model parameters is whether or not to use the sampling weights. Survey samplers are quite accustomed to using survey weights to estimate means, totals, ratios, proportions, etc. For example,

$$\hat{\bar{Y}} = \Sigma\, w_i\, y_i \;/\; \Sigma\, w_i$$

and

$$\hat{R} = \Sigma\, w_i\, y_i \;/\; \Sigma\, w_i\, x_i$$

are commonly used estimators of means and ratios. More complex estimators of these quantities, which are model-based but design-dependent, may be found in quite general terms in Sarndal (1982).

However, for more complex parameters such as regression coefficients, the considerations for whether to use the sampling weights are more subtle. The answer seems to depend on the following points:

(a) What model is being fitted?

(b) What are the parameters of interest?

(c) How much faith can be placed on the universal applicability of the model?

(d) How large is the sample size?

Even for the common case of fitting regression coefficients, a number of models have been proposed in the literature. DuMouchel and Duncan (1983) study four models for stratified random sampling. These are:

I. $Y = X\beta + \varepsilon$ (2.1)

where the conditional distribution of the ε's (conditional on the stratum identification) is such that the ε's are uncorrelated, homoscedastic random variables with zero mean. Note that β is constant over all strata.

II. $Y = X\beta_j + \varepsilon$ (2.2)

for stratum j, where the ε's have the same properties as in I.

III. $Y = X\beta + U\gamma + \varepsilon$ (2.3)

where U is not, in general, observable, but has been transformed so as to be orthogonal to X; i.e. $E(X^T U) = 0$. Again the ε's behave as in I.

IV. $Y = X\beta^* + \varepsilon^*$ (2.4)

where it is simply assumed that the ε^*'s and the X-matrix are uncorrelated and the ε^*'s have mean zero.

Often, Models I, II and III are called superpopulation models. Holt, Smith and Winter (1980) say, in a more general context, that although Model I may be true before sample selection, if after sample selection $E(\varepsilon|X) \neq 0$ then the sample design is informative.

Porter (1973), Pfefferman and Nathan (1977) and others propose a random coefficient regression model

$$Y_i = X_i \beta_i + \varepsilon_i \tag{2.5}$$

where a certain structure on the β_i's is imposed.

Many authors emphasize that any models implied by the analysis must be consistent with the data; see, for example Brewer and Mellor (1973), Nathan (1981), DuMouchel and Duncan (1983) and Kalton (1983).

Under Model I,

$$\begin{aligned} B &= [E(X'X)]^{-1} E(X'Y) \\ &= \beta + O_p(N^{-\frac{1}{2}}), \end{aligned} \tag{2.6}$$

where the expectation operator here refers to the finite population value. Therefore, for large finite populations, the distinction between β and B can be virtually ignored. Binder (1983) extended this analogy to parameters of Generalized Linear Models and other forms of implicitly defined parameter. For Model II, DuMouchel and Duncan (1983) consider the parameter of interest to be $\Sigma W_j \beta_j$ where the W's are proportional to the stratum sizes and $\Sigma W_j = 1$. This could be extended to general W's as in Pfefferman and Nathan (1977). In Model III the parameter of interest is β. Holt, Smith and Winter (1980) consider a model analogous to Model III, where the U's are observable (see Section 3.4). There, the parameter of interest is $V_{xy} V_{xx}^{-1}$ where V refers to the superpopulation covariance matrix. In Model IV, the parameter of interest is B given by (2.6).

We see that if the classical model (2.1) holds universally, then the best linear unbiased estimator for β is the usual ordinary least squares estimator, ignoring the sample design information and the sample weights. Nathan (1981) raises questions as to the relevance of B is when the model does not hold. A partial answer to this quandary is given by studying Model III. First, one should ensure that the model is approximately true. This may include adding design variables to the model to reduce the correlation of the residuals. Secondly, the model may still be useful for predictive purposes, if the missing variables behave similarly at the prediction point. In this way, the analyst can view the regression coefficients as being descriptive population parameters. For example, Freeman, Freeman and Ostfeld (1983) use the regression formulation to make an age-sex-race adjustment to their dependent variable, in order to facilitate comparisons at two different time points.

If Model I is inappropriate, but a formulation similar to Model III is appropriate, then the weighted regression coefficient estimator (using the sampling weights) will usually provide a consistent estimator of B. Even if Model I is true, the use of the sampling weights will lead to unbiased but inefficient estimates of β. This may not be serious to the practitioner if the sample size is large. Nathan (1981), DuMouchel and Duncan (1983) and Fuller (1984) have suggested fitting models which allow for testing whether the weighted and unweighted estimates of the regression coefficients have the same expected value. Unfortunately, unless the weights are very different, this test will not be very powerful because the assumed models tend to have a high degree of multi-collinearity among the independent variables.

All of the above discussions have assumed large populations with large samples. If the sample size is small, the inefficiency resulting from using the weighted estimates may be more serious when the model is in fact true. Also, if the model is only approximately true, the variance reduction resulting from using the unweighted estimates may more than compensate for the increase in bias, thus making the unweighted estimates the preferred choice. In this case, the search for all the relevant variables of the model becomes even more important.

3.3 Making Inferences

The use of the sample design information when developing confidence intervals or performing tests of hypotheses can be crucial. If the analyst has complete faith in a model such as Model I, clearly it is appropriate to apply the classical methods available for such models (including appropriate robust methods available for such models for long-tailed error distributions). However, the danger of applying such methods when the model is not strictly true has been demonstrated by Scott and Holt (1982), where a block diagonal correlated error structure can introduce serious biases into the variance estimates, even for self-weighting designs.

Also, the distinction between a superpopulation parameter, such as β, and its finite population analogue, B, is important if the sampling fractions are large. For example, in a one-way analysis of variance set-up, with two categories it may be more appropriate to test the hypothesis

H_o: $|\bar{Y}_2 - \bar{Y}_1| \leq \delta$ for a suitable δ, rather than

H_o: $\bar{Y}_2 - \bar{Y}_1 = 0$; see Nathan (1981).

Variance estimation for parameters in a design-based setting is usually performed by balanced repeated replication, jackknifing or Taylor linearization.

Taylor linearization for explicitly defined parameters is given in Tepping (1968). Woodruff (1971) showed a simpler method for computing the variances by taking total differentials of the functions to be estimated. Woodruff and Causey (1976) describe a computer program for calculating these variances, where the derivatives are numerically computed.

For the case of estimating regression coefficients, Fuller (1975) gave a simpler computational formula than that given by Tepping (1968). Binder (1983) extended this to more complex situations of implicitly defined parameters; for example, parameters of Generalized Linear Models and M-estimators. Shah, Holt and Folsom (1977), and Lemeshow and Stoddard (1984) provide simulations demonstrating that the nominal significance levels for regression coefficients using Taylor linearization perform quite well even when the number of strata and the number of observations per stratum are small.

Another method used for variance estimation is the Quenouille - Tukey jackknife (Quennouille, 1956; Tukey, 1958) which was introduced in the late fifties and discussed and developed in the context of simple random sampling or sampling from infinite populations by Brillinger (1964), Miller (1964, 1968, 1974a, 1974b) and Gray and Schucany (1972) amongst others. The bootstrap method, on the other hand, was introduced relatively recently by Efron (1979) and compared to the jackknife and other resampling plans in Efron (1982) for the same class of problems.

McCarthy (1969), Folsom, Bayless and Shah (1971), Frankel (1971), Jones (1974) and Sharot (1976) were among the first to attempt to reuse the sample by jackknifing in addition to balanced repeated replication (Kish and Frankel, 1970) in the context of finite sampling. Their studies are empirical in nature, considering for the most part, stratified (without replacement) designs. The results support earlier observations of the usefulness of the jackknife and balanced repeated replication for the purpose of variance estimation and bias reduction. The parameters of concern are functions of stratum means, ratios and correlation coefficients being the most common, although weighted estimates of regression coefficients would equally apply. Further empirical extensions can be found in Brillinger (1966), Kish and Frankel (1974), Bean (1975) and Lemeshow and Epp (1977), while Wolter (1979) provides a useful summary and overview.

By contrast, while the use of the bootstrap technique in the context of finite sampling is not uncommon (although infinite populations are generally assumed), the study of its properties is surprisingly rare. McCarthy and Snowden (1983) provide one of the few empirical investigations of the method as it applies to finite populations. They consider two approaches. The first, suggested by Bickel and Freedman (1984), consists of creating a superpopulation from the sample, by replicating each of its units a number of times equal to its weight. The superpopulation is then sampled without replacement. McCarthy and Snowden, on the other hand, resample the sample with replacement, choosing the bootstrap sample size such that the variance of the bootstrap estimator of the mean is correct. Subsequent studies suggest,

however, that this method fails to capture moments of higher order and can thus be misleading if applied to highly skewed populations.

Theoretical evaluation of sample reuse methods in the context of finite sampling has become of interest only recently and is almost exclusively confined to the study of asymptotic behaviour. Krewski (1978b) provides some insight as to the stability and efficiency of the balanced repeated replication variance estimator, as a function of the number of subsamples selected, for the case of two or more units per stratum; while Royal and Cumberland (1978) justify asymptotically the jackknife variance estimator of the population total, demonstrating favourable results as compared to the best linear unbiased estimator under a linear regression model. The key results however, were put forth by Krewski (1978a), where he proves asymptotic normality and the consistency of the jackknife variance estimator for functions of U-statistics in the case of simple random sampling with replacement. Krewski's results were extended by Majumdar and Sen (1978) who prove the strong laws for the jackknife variance estimator of U-statistics. Krewski and Rao (1981) later provide some first order asymptotics in the case of stratified (with replacement) designs for both the balanced repeated replication and the jackknife variance estimators for functions of stratum means. They consider the case where the number of strata tends to infinity rather than the simpler case of large stratum sizes. Rao and Wu (1983b) provide the second order properties for this problem. A good overview of the jackknife and balanced repeated replication asymptotics is given in Rust (1984).

The asymptotic theory for the bootstrap for finite samples is scarce. Babu and Singh (1983) prove some asymptotic properties of the bootstrap variance estimator by letting the stratum sample size go to infinity. Bickel and Freedman (1984) generalize these results by considering stratified samples selected with or without replacement and by letting the total sample size tend to infinity (that is, either the number of strata, or the stratum sample sizes or both tend to infinity). They show that the bootstrap (in their superpopulation context) and the jackknife variance estimators are asymptotically normal and consistent for the case of estimating linear combinations of stratum means. Rao and Wu (1983a) extend these first order asymptotic results for the bootstrap variance estimator to various other designs including unequal probability designs. Their method of bootstrapping, however, involves the use of adjusted values, somewhat analogous to Tukey's pseudo-values, for the jackknife.

All of the above procedures rely on asymptotics for their theoretical justification. Fuller (1984) gives the following rules of thumb for the applicability of large sample theory:

1. The mean of the Taylor deviates must be approximately normal (i.e. no influential outliers).
2. There must be sufficient effective degrees of freedom to estimate the variances well.
3. The curvature of the non-linear function of sample means or the standard error of the means must be small.

3.4 An Alternative Approach to Regression

In Section 3.2 we discussed four regression models, specific to stratified sample designs. Here, another model, proposed in Holt, Smith

and Winter (1980) is discussed. The model, with its corresponding strategy is fundamentally different in that it is a maximum likelihood technique (using a superpopulation model) that also purports to account for the sample design. It does this, without using the sampling weights, by incorporating "design variables" in a linear superpopulation model.

The finite population is viewed as a random sample of size N from a superpopulation. The vector of variables, X, can be partitioned into

X_1: the dependent variables

X_2: the independent, carrier or explanatory variables

and X_3: the design variables.

The variables are assumed to have a multivariate normal distribution in the superpopulation (this assumption can be relaxed to a set of second order conditions, as in Nathan and Holt; 1980). We let μ_i be $E(X_i)$ and V_{ij} be $E((X_i - \mu_i)(X_j - \mu_j)^T)$. The value of the design variable vector X_3, is known for all N units in the finite population, but X_1, and X_2 are only known for the n units in the sample.

The parameter of interest (β) is the best linear predictor of X_1 as a function of X_2, given by $V_{12}\ V_{22}^{-1}$.

If X_1 and X_2 were known for all the units in the population we could estimate V_{12} and V_{22}^{-1} in the usual way to give us $\hat{\beta}_1 = \hat{V}_{12}\ (\hat{V}_{22})^{-1}$. If X_3 was only known for the sampled units then the maximum likelihood estimate of β would be the usual ordinary least squares (OLS) estimator $\hat{\beta}_2 = S_{12}\ (S_{22})^{-1}$, where S_{ij} is the sample covariance matrix, based on the n sampled observations.

However the dataset consists of n complete sampled observations, and N-n incomplete observations consisting of those units for which only the design information is known. The maximum likelihood estimator for this problem has been discovered by a number of authors (for instance, DeMets and Halperin; 1977):

$$\hat{\beta} = (S_{12} + S_{13} S_{33}^{-1} (\hat{V}_{33} S_{33}^{-1} - I) S_{32}) \\ (S_{22} + S_{23} S_{33}^{-1} (\hat{V}_{33} S_{33}^{-1} - I) S_{32})^{-1}, \qquad (3.1)$$

where $\hat{V}_{33}$ is the covariance matrix for X_3 based on the N population values.

The OLS estimate ($\hat{\beta}_2$) will be the same as $\hat{\beta}$ when $S_{32} = 0$ or $\hat{V}_{33} = S_{33}$. If S_{32} is zero, then there is no observed correlation between X_3 and X_2; that is, we observed no linear relationship between the design variables and the explanatory variables. When the survey is a simple random sample or at least approximately self-weighting, $\hat{V}_{33}$ and S_{33} will usually be approximately equal. However, if there are wide variations in the probabilities of selection and joint inclusion probabilities, S_{33} can be expected to be quite different from $\hat{V}_{33}$. Hence, while the weights are not used in the estimation of $\hat{\beta}$, they will to a large extent determine whether this estimate is different from the OLS estimate $\hat{\beta}_2$. Pfefferman and Holmes (1985) demonstrate the possible biases of the procedure under certain types of model failure.

Yet another alternative would be to use the estimator for $\hat{\beta}$ in (3.1) based on $\hat{S}_{ij}$, the weighted sample covariance matrix. This would provide design-consistent estimates of $V_{12} V_{22}^{-1}$.

4. CATEGORICAL DATA ANALYSIS

In this section we review survey data anlysis, particularly as it relates to analyzing categorical data. This is an important area of application since, for many surveys, responses in categories are easier to obtain than quantitative variables. Techniques for fitting various log-linear and logistic regression models and undertaking the related analyses are described in Bishop, Fienberg and Holland (1975), and also have been the subject of excellent review articles by Imrey, Koch and Stokes (1981, 1982). The treatment in Bishop, Fienberg and Holland is restricted to the case of cell totals following a Poisson, multinomial or product multinomial sampling distribution. Grizzle, Starmer and Koch (1969) proposed a general approach for analyzing multivariate catagorical data by using linear models. Their methodology, based theoretically on the work of Wald (1943), specified procedures for fitting linear models to functions of the unknown true cell probabilities. Subsequent work by Koch, Freeman and Freeman (1975) and Schuster and Downing (1976), based on methods similar to those of Grizzle, Starmer and Koch, is intended for data obtained from complex sampling designs. We give a brief summary of their methods.

Let i = 1, 2, ..., s index the set of distinct sub-populations from which samples are drawn and j=1, 2, ..., r index the set of response profiles. Denoting the true proportion of the ith sub-population and the jth response category by π_{ij}, we define

$$\underset{\sim}{\pi}^T = (\pi_{11}, \ldots, \pi_{1r}, \pi_{21}, \ldots \pi_{2r}, \ldots, \pi_{s1}, \ldots, \pi_{sr})$$

with $\sum_j \pi_{ij} = 1$ for each i. Suppose that a sample from each sub-population is drawn according to a specified survey design and

$$\underset{\sim}{p}^T = (p_{11}, \ldots, p_{1r}, p_{21}, \ldots, p_{2r}, \ldots, p_{s1}, \ldots, p_{sr})$$

is a design-based estimate of $\underset{\sim}{\pi}^T$. We are interested in functions $F_i(\underset{\sim}{\pi})$, (i=1, ... u) of $\underset{\sim}{\pi}$. These functions define the relations of interest between the response categories and sub-populations. Let

$$F(\underset{\sim}{\pi})^T = (F_1(\underset{\sim}{\pi}), F_2(\underset{\sim}{\pi}), \ldots, F_u(\underset{\sim}{\pi})).$$

We assume that $F_i(\underset{\sim}{\pi})$'s are functionally independent so that the covariance matrix of the estimate of $F(\underset{\sim}{\pi})$ is non-singular. The variation among the u elements of $F(\underset{\sim}{\pi})$ may be investigated by fitting the model

$$M: \quad F(\underset{\sim}{\pi}) = X\,\beta$$

where X is a design matrix of rank $v \leq u$ and β is a v x 1 matrix of unknown parameters. For some problems, (e.g. homogeneity of margins) the model of interest may be $F(\underset{\sim}{\pi}) = 0$.

One of several methods of estimation, such as the maximum likelihood, ordinary least squares, weighted least square (WLS), and minimum chisquare may be used for estimating β. The choice of method may depend on the statistical properties of the estimates and the computational efficiency of the associated algorithms. Koch and his co-workers advocate the use of Weighted Least Squares (WLS) methods. In that case

$$b = (X^T V_F^{-1} X)^{-1} \; X^T V_F^{-1} F,$$

Where V_F is a design-based consistent estimate of the covariance matrix of $F = F(\underset{\sim}{p})$. Note that the construction of the Wald statistic and the WLS estimate of b depend on the ability to get the estimate V_F. Goodness of fit tests are possible by defining the saturated model as $F(\underset{\sim}{\pi}) = X_1 \beta_1 + X_2 \beta_2$

and testing for $\beta_2 = 0$. In general, when the model has been found consistent with the data, statistical tests concerning linear hypotheses involving β may also be carried out. For testing H: $C\beta = 0$ where C is a known (d x u) matrix, we use the test statistic

$$Q_C = b^T C^T [C V_b C^T]^{-1} Cb$$

where V_b is a consistent estimate of the covariance matrix of b. The statistic Q_C has an asymptotic χ^2 distribution with d d.f. under the hypothesis H. These results are easily extended to other estimates besides the WLS estimate; see Binder (1983) for the appropriate variance estimates. His general method includes estimating Generalized Linear Model parameters as described in Nelder and Wedderburn (1972).

The weighted least squares methodology, developed by Koch, Freeman and Freeman (1975), has provided a useful and unified approach to the problem of analyzing data from complex survey designs. However, its use is limited in some applications by the necessity to produce an estimate of the covariance matrix of F. Difficulties could arise from the required inversion of V_F when the number of cells is large or if the number of observations in some cells is small. Detailed discussion of the limitation of WLS methodology is given in Fay (1984, 1985).

Nathan (1969, 1975) has also developed asymptotic methods for testing independence in contingency tables from stratified samples. His method, based on maximum likelihod estimates also requires estimation of the covariance matrix of the cell estimates.

For data collected by simple random samples, methods for testing the goodness of fit of the model, using the Pearson's chi-squared (X^2) and

likelihood ratio (G^2) statistics are well-known. Their use has been considerably enhanced by availability of standard computer packages. However, in the context of the complex survey design, these statistics have a serious deficiency. Unlike the Wald Statistic, X^2 (or G^2) does not usually have an asymptotic χ^2 distribution under the model of interest when clustering and stratification have been used in designing the survey. Consequently the conclusions based on these statistics may not be valid. Fellegi (1980), Rao and Scott (1981) and others have shown that clustering and stratification can have a considerable effect on the distribution of X^2 and hence its use (without any adjustment) can give misleading results in practice.

Recent work has been directed at investigating the asymptotic distribution of X^2 and G^2 under cluster sampling and other multi-stage sampling designs and adjustments to X^2 have been proposed. Cohen (1976), Altham (1976), Choi (1981) and Brier (1980) have investigated the distribution by modelling probabilities for cluster sampling. They have shown that, under several hypotheses, X^2 divided by a suitable constant has an asymptotic χ^2 distribution. However, their methods impose a constraint of equal design effects of the estimates. This constraint is rarely satisfied in practice. Fellegi (1980) proposed the use of $X^2/d.$ (as having χ^2 distribution) as a test statistic, where d. is the average of the design effects of the cell estimates. The computation of d. requires less than full knowledge about the covariance structure of sample estimates than that required by the Wald statistic.

The asymptotic distribution of X^2, based on data from a complex survey design, has been investigated by Rao and Scott (1981, 1984) and

Roberts (1985). It is shown that, for a wide class of models, the appropriate chi-squared (X^2) test statistic is asymptotically distributed as $\sum_{i=1}^{k} \delta_i Z_i$, the weighted sum of independent χ^2 variables Z_i. Each Z_i has 1 degreee of freedom and k depends on the hypothesis. The weights δ_i are the eigenvalues of the matrix $A_o^{-1} A_1$ where A_1 is the design based covariance matrix of a statistic used in defining X^2 and A_o is the corresponding matrix under multinomial sampling. The quantitities δ_i are called the generalized design effects and are consistently estimated by the eigenvalues of $\hat{A}_o \hat{A}_1^{-1}$, where $\hat{A}_o$ and $\hat{A}_1$ are consistent estimates of A_o and A_1, respectively. We denote the estimate of δ_i by $\hat{\delta}_i$.

Due to the non-standard nature of the distribution of X^2, it is convenient to approximate it by some standard distribution function. A first order approximation to the distribution of X^2 can be obtained by treating $X^2 = k\hat{\delta}.\chi^2_k$ where $\hat{\delta}. = \sum_{i=1}^{k} \hat{\delta}_i / k$ and k is the number of positive $\hat{\delta}_i$'s. Hence we may use the adjusted test statistic $X^2_C = X^2/\hat{\delta}.$ and regard it has having a χ^2_k distribution. The adjusted test statistic X^2_C (or G^2_C) provides a satisfactory correction except for the case where $\{\delta_i\}$ has a large coefficient of variation. This adjustment, like the one proposed by Fellegi (1980), requires less than full knowledge about the covariance structure of the estimates.

A better adjustment, based on Satterthwaite's approximation, treats $X^2_S = X^2_C/(1 + a^2)$ as having χ^2 distribution with v degrees of freedom where

$$v = (k - 1) / (1 + a^2)$$

and $$a^2 = \sum_i (\hat{\delta}_i - \hat{\delta}.)^2 / [(k-1)\,\hat{\delta}.]^2$$

Note that a^2 is the square of the coefficient of variation of the $\hat{\delta}_i$'s. This approximation is useful when a^2 is significantly different than zero. However, this approximation generally requires full knowledge of the covariance matrix. These methods have been used extensively by Hidiroglou and Rao (1983) for analysis of Canada Health Survey data. Comparative results on the significance level due to the use of X^2, $X^2/d.$, X_C^2 and X_S^2 are also provided.

The analysis of survey data using logit and other transformation models has been investigated by Roberts (1985). The methodology proposed is similar to that of Rao and Scott (1984) and has been used by Kumar and Rao (1984, 1985) for analyzing data from the Canadian Labour Force Survey. Kumar and Rao (1984) have also developed diagnostic procedures, under a logit model based on survey data, for detecting outliers and influential points.

Fay (1985) proposes another kind of adjustment to the standard Pearson and likelihood ratio chi-squared statistics. His approach is based on jackknifing these statistics. He proposes a new test (based on replication) called the "jackknife chi-square test" and is used for fitting a log-linear model for survey data. The original cross-classified data is split into a series of replicates. A good discussion on the choice of replication strategies is given in Fay (1983). For each replicate, X^2 (or G^2) is estimated through the solution of maximum likelihood equations appropriate to simple random sampling. The jackknife chi-squared statistics, based on these estimates, is computed. A computer program, called CPLX, for implementing jackknifed chi-square test has also been developed by Fay (1983). The program permits three types of replication methods: the simple jackknife, the stratified jackknife and half sample methods. The choice of

the replicate methods is left to the users. This method has been applied to a number of analyses done at the U.S. Bureau of Census. No comparison of the jackknife chi-squared test and those proposed by Rao and Scott has yet been conducted.

6. PERCENTILE ESTIMATION

Even for the 'standard' case of infinite population sampling, estimation of the median and other percentiles, along with associated standard errors and confidence intervals, is not an easy problem. Various alternatives include the sample median, weighted linear functions of the order statistics (Harrell and Davis; 1982), estimates based on kernel density estimates (Sheather and Maritz; 1983) and others. Standard errors of the sample median using bootstrapping (Efron; 1979, 1982 and Ghosh, et al.; 1984) or using other "smoothed" estimates (Sheather and Maritz; 1983) are available.

For survey data collected from finite populations, Woodruff (1952) described a technique for obtaining confidence intervals under general sampling plan, by projecting the confidence interval for the binomial proportion onto the estimated distribution function. Exact intervals under simple random sampling are given in Wilks (1962) and Konijn (1973). McCarthy (1965) looked at some exact intervals under stratified sampling with proportional allocation. Sedransk and Meyer (1978) obtained exact intervals under stratified sampling in general. Gross (1980) proposed a bootstrap variance estimate for stratified samples. Bayesian methods are given in Hill (1968) and Binder (1982). The latter author obtained intervals that were asymptotically equivalent to Woodruff's (1952) intervals.

This issue is clearly one that requires much more theoretical and empirical research.

7. SUMMARY

We have concentrated here on the importance of the sample design when analyzing survey data. Although, with the wide availability of commercially available software, it is possible to perform quite sophisticated data analyses, these approaches are often inapplicable to data collected under complex survey designs. We have touched on a wide variety of problems, such as model parameter estimation and tests of significance. However, there are still many unresolved problems with respect to incorporating the design considerations when analyzing multivariate data from complex designs. Examples include cluster analysis, factor analysis, time series methods, multi-dimensional scaling and so on. As the users of survey data become more sophisticated and there is more awareness in the statistical community of the inherent problems, research into these areas will flourish.

BIBLIOGRAPHY

Altham, P.M.E. (1976). Discrete variable analysis for individuals grouped into families. Biometrika, 63, 263-269.

Babu, G.J. and Singh, K. (1983). Inference on means using the bootstrap. Ann. Statist., 11, 999-1003.

Bean, J.A. (1975). Distribution and properties of variance estimators for complex multistage probability samples. Vital and Health Statistics Series 2, No. 65.

Bickel, P.J. and Freedman, D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist., 12, 470-482.

Binder, D.A. (1982). Non-parametric Bayesian models for samples from finite populations. J.R. Statist. Soc. B, 44, 388-393.

Binder, D.A. (1983). On the variances of asymptotically normal estimators from complex surveys. Intl. Statist. Rev., 51, 279-292.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975). Discrete Multivariate Analysis, MIT Press, Cambridge.

Brewer, K.R.W. and Mellor, R.W. (1973). The effect of sample structure on analytical surveys. Austral. J. Statist., 15, 145-152.

Brier, S.E. (1980). Analysis of contingency table data under cluster sampling. Biometrika, 67, 591-596.

Brillinger, D.R. (1964). The asymptotic behaviour of Tukey's general method of setting approximate confidence limits (the jackknife) when applied to maximum likelihood estimates. Intl. Statist. Rev., 32, 202-206.

Brillinger, D.R. (1966). The application of the jackknife to the analysis of sample surveys. Commentary, 8, 74-80.

Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1979), Prediction theory for finite populations when model-based and design-based principles are combined. Scand. J. Statist., 6, 97-106.

Choi, J.W. (1981). A further study on the analysis of categorical data from weighted cluster sample survey. Proc. Am. Statist. Assoc., Section on Survey Methods Research, 15-20.

Cochran, W.G. (1953). Sampling Techniques, Wiley, New York.

Cohen, J.E. (1976). The distribution of the chi-squared statistics under cluster sampling from contingency tables. J. Am. Statist. Assoc., 71, 665-670.

DeMets, D. and Halperin, M. (1977). Estimation of simple regression coefficients in samples arising from sub-sampling procedures. Biometrics, 33, 47-56.

Deming, W.E. (1950). Some Theory of Sampling. Wiley, New York.

DuMouchel, W.H. and Duncan, G.J. (1983). Using sample survey weights in multiple regression analysis of stratified samples. J. Am. Statist. Assoc., 78, 535-543.

Efron. B. (1979). Bootstrap methods: Another look at the jackknife. Ann. Statist., 7, 1-26.

Efron. B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans. SIAM, Philadelphia.

Fay, R.E. (1982). Contingency tables for complex designs: CPLX. Proc. Am. Statist. Assoc., Section on Survey Research Methods, 44-53.

Fay, R.E. (1983). CPLX - Contingency Tables Analysis for Complex Sample Designs, Program Documentation. Unpublished report. U.S. Bureau of the Census, Washington.

Fay, R.E. (1984). Some properties of estimates of variance based on replication methods. Proc. Am. Statist. Assoc., Section on Survey Research Methods.

Fay, R.E. (1985). A jackknifed chi-square test for complex samples. J. Am. Statist. Assoc., 80, 148-157.

Fellegi, I.P. (1980). Approximate tests of independence and goodness of fit based on stratified multistage samples. J. Am. Statist. Assoc., 75, 261-268.

Folsom, R., Bayless, D.L. and Shah, B.V. (1971). Jackknifing for variance components in complex sample survey designs. Invited paper at the American Statistical Association Meetings, Fort Collins, Colorado.

Frankel, M.R. (1971). Inference from Survey Samples. Ann Arbor, Michigan.

Freeman, J.L., Freeman, D.H. Jr. and Ostfeld, A.M. (1983). Analysis of data from successive complex sample surveys, with an example of hypertension prevalence from the United States Health Examination Survey. Intl. J. Epidem., 12, 230-237.

Fuller, W.A. (1975). Regression analysis for sample survey. Sankhya C, 37, 117-132.

Fuller, W.A. (1984). Least squares and related analyses for complex survey designs. Survey Methodology, 10, 97-118.

Ghosh, M., Parr, W.C., Singh, K. and Babu, G.J. (1984). A note on bootstrapping the sample median. Ann. Statist. 12, 1130-1135.

Gray, H.L. and Schucany, W.R. (1972). The Generalized Jackknife Statistic. Dekker, New York.

Grizzle, J.E., Starmer, C.F. and Koch, G.G. (1969). Analysis of categorical data by linear models. Biometrics, 25, 489-504.

Gross, S.T. (1980). Median estimation in sample surveys. Proc. Am. Statist. Assoc., Section on Survey Research Methods, 181-184.

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. (1953). Sample Survey Methods and Theory, Vol. I and II. Wiley, New York.

Harrell, F.E. and Davis, C.E. (1982). A new distribution-free quantile estimator. Biometrika, 69, 635-640.

Hidiroglou, M.A. and Rao, J.N.K. (1983). Chi-squared test for the analysis of three way contingency tables from the Canada Health Survey. Technical report. Statistics Canada.

Hill, B.M. (1968). Posterior distribution of percentiles: Bayes' theorem for sampling from a population. J. Am. Statist. Assoc., 63, 677-691.

Holt, D., Smith, T.M.F. and Winter, P.D. (1980). Regression analysis of data from complex surveys. J.R. Statist. Soc. A, 143, 474-487.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. J. Am. Statist. Assoc., 47, 663-685.

Imrey, P.B., Koch, G.G. and Stokes, M.E. (1981, 1982). Categorical data analysis: Some reflections on the log linear model and logistic regression: Parts I and II. Intl. Statist. Rev., 49, 265-283 and 50, 35-64 (in collaboration with J.N. Darroch, D.H. Freeman Jr. and H.D. Tolley).

Imrey, P.B., Sobel, E. and Francis, M. (1980). Modelling contingency tables from complex surveys. Proc. Am. Statist. Assoc., Section on Survey Research Methods, 212-217.

Jones, H.L. (1974). Jackknife estimation of functions of stratum means. Biometrika, 61, 343-348.

Kalton, G. (1983). Models in the practice of survey sampling. Intl. Statist. Rev., 51, 175-188.

Kish, L. and Frankel, M.R. (1970). Balanced repeated replications for standard errors. J. Am. Statist. Assoc., 65, 1071-1094.

Kish, L. and Frankel, M.R. (1974). Inference from complex samples. J.R. Statist. Soc. B, 36, 1-37.

Klein, L.R. and Morgan, J.N. (1951). Results of alternative statistical treatments of sample survey data. J. Am. Statist. Assoc., 46, 442-460.

Koch, G.G., Freeman, D.H. Jr. and Freeman, J.L. (1975). Strategies in the multivariate analysis of data from complex surveys. Intl. Statist. Rev., 43, 59-78.

Konijn, H.S. (1962). Regression analysis in sample surveys. J. Am. Statist. Assoc., 57, 590-606.

Konijn, H.S. (1973). Statistical Theory of Sample Survey Design and Analysis. North Holland, Amsterdam.

Krewski, D. (1978a). Jackknifing U-statistics in finite populations. Comm. Statist., A7, 1-12.

Krewski, D. (1978b). On the stability of some replication variance estimators in the linear case. J. Statist. Planning and Inference, 2, 45-51.

Krewski, D. and Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife and balanced repeated replication methods. Ann Statist., 9, 1010-1019.

Kumar, S. and Rao, J.N.K. (1984). Logistic regression analysis of Labour Force Survey data. Survey Methodology, 10, 62-81.

Kumar, S. and Rao, J.N.K. (1985). Fitting Box-Cox transformation models to Labour Force Survey data. Survey Methodology, to appear.

Lemeshow, S. and Epp, R. (1977). Properties of the balanced half-sample and jackknife variance estimation techniques in the linear case. Comm. Statist., A6, 1259-1274.

Lemeshow, S. and Stoddard, A.M. (1984). A comparison of alternative variance estimation strategies for estimating the slope of a linear regression in sample surveys. Comm. Statist. - Simula. Computa., 13, 153-168.

Majumdar, H. and Sen, P.K. (1978). Invariance principles for jackknifing U-Statistics for finite population sampling and some applications. Comm. Statist., A7, 1007-1025.

McCarthy, P.J. (1965). Stratified sampling and distribution-free confidence intervals for a median. J. Am. Statist. Assoc., 60, 772-783.

McCarthy, P.J. (1969). Pseudo-replication: Half samples. Intl. Statist. Rev., 37, 239-264.

McCarthy, P.J. and Snowden, C.B. (1983). The bootstrap and finite population sampling. Unpublished manuscript.

Miller, R.G. (1964). A trustworthy jackknife. Ann. Math. Statist., 35, 1594-1605.

Miller, R.G. (1968). Jackknifing variances. Ann. Math. Statist., 39, 567-582.

Miller, R.G. (1974a). An unbalanced jackknife. Ann Statist., 2, 880-891.

Miller, R.G. (1974b). The jackknife - a review. Biometrika, 61, 1-17.

Nathan, G. (1969). Tests of independence in contingency tables from stratified samples. New Developments in Survey Sampling, 578-600. (N.L. Johnson and H. Smith, eds.). Wiley, New York.

Nathan, G. (1975). Tables of independence in contingency tables from stratified samples. Sankhya C, 37, 77-87.

Nathan, G. (1981). Notes on inference based on data from complex sample designs. Survey Methodology, 7, 77-87.

Nathan, G. and Holt, D. (1980). The effect of survey design on regression analysis. J.R. Statist. Soc. B, 42, 377-386.

Nelder, J.A. and Wedderburn, R.W.M. (1972). Generalized linear models. J.R. Statist. Soc. A., 135, 370-384.

Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. J.R. Statist. Soc., 97, 558-625.

Porter, R.D. (1973). On the use of survey sampling weights in the linear model. Ann. Econ. Soc. Meas., 2, 141-158.

Pfefferman; D. and Nathan, G. (1977). Regression analysis of data from complex samples. Bull. Intl. Statist. Inst., 41(3), 21-42.

Quenouille, M.H. (1956). Notes on bias in estimation. Biometrika, 43, 353-360.

Rao, J.N.K. and Hidiroglou, M.A. (1981). Chi-squared tests for the analysis of categorical data from the Canada Health Survey. Bull. Intl. Statist. Inst., 49(2), 699-718.

Rao, J.N.K. and Scott, A.J. (1979). Chi-squared tests for analysis of categorical data from complex surveys. Proc. Am. Statist. Assoc., Section on Survey Research Methods, 58-66.

Rao, J.N.K. and Scott, A.J. (1981). The analysis of categorical data from complex surveys: Chi-squared tests for goodness of fit and independence in two way tables. J. Am. Statist. Assoc., 76, 221-230.

Rao, J.N.K. and Scott, A.J. (1984). On chi-square tests for multiway contingency tables will cell proportions estimated from survey data. Ann. Statist., 12, 46-60.

Rao, J.N.K. and Wu, C.F.J. (1983a). Bootstrap inference with stratified samples. Laboratory for Research in Statistics and Probability, Technical Report No. 19.

Rao, J.N.K. and Wu, C.F.J. (1983b). Inference from stratified samples: Second order analysis of three methods for non-linear statistics. Laboratory for Research in Statistics and Probability, Technical Report No. 7.

Roberts, G.R. (1985). Contributions to chi-squared tests with survey data. Carleton University. Ph. D. Thesis.

Royall, R.M. and Cumberland, W.G. (1978). Variance estimation in finite population sampling. J. Am. Statist. Assoc., 73, 351-358.

Rust, K.F. (1984). Techniques for estimating variances for sample surveys. University of Michigan. Ph.D. Thesis.

Särndal, C.E. (1982). Implications of survey design for generalized regression estimation of linear functions. J. Statist. Planning and Inference, 7, 156-170.

Schuster, J.J. and Downing, D.J. (1976). Two-way contingency tables for complex sampling schemes. Biometrika, 63, 271-276.

Scott, A.J. and Holt, D. (1982). The effect of two-stage sampling on ordinary least squares methods. J. Am. Statist. Assoc., 77, 848-854.

Scott, A.J. and Rao, J.N.K. (1981). Chi-squared tests for contingency tables with proportions estimated from survey data. Current Topics in Survey Sampling (D. Krewski, R. Platek and J.N.K. Rao, eds.).

Sedransk, J. and Meyer, J. (1978). Confidence intervals for the quantiles of a finite population: Simple random and stratified simple random sampling. J.R. Statist. Soc. B, 40, 239-252.

Shah, B.V., Holt, M.M. and Folsom, R.E. (1977). Inference about regression models from sample survey data. Bull. Intl. Statist. Inst., 41(3), 43-57.

Sharot, T. (1976). The generalized jackknife: Finite samples and subsample sizes. J. Amer. Statist. Assoc., 71, 451-454.

Sheather, S.J. and Maritz, J.S. (1983). An estimate of the asymptotic standard error of the sample median. Austral. J. Statist., 25, 109-122.

Sukhatme, P.V. (1953). Sampling Theory for Surveys with Applications, Iowa State.

Tepping, B.J. (1968). Variance estimation in complex surveys. Proc. Am. Statist. Assoc., Social Statistics Section, 11-18.

Tukey, J.W. (1958). Bias and confidence in not-quite large samples (Abstract). Ann. Math. Statist., 29, 614.

Wald, A. (1943). Tests of statistical hypothesis concerning general parameters when the number of observations is large. Trans. Am. Math. Soc., 54, 426-482.

Wilks, S.S (1962). Mathematical Statistics, 2nd ed. Wiley, New York.

Wolter, K.M. (1979). Variance Estimation. U.S. Bureau of the Census, Course Notes.

Woodruff, R.S. (1952). Confidence intervals for medians and other position measures. J. Am. Statist. Assoc., 47, 635-646.

Woodruff, R.S. (1971). A simple method for approximating the variance of a complicated estimate. J. Am. Statist. Assoc., 66, 411-414.

Woodruff, R.S. and Causey, B.D. (1976). Computerized method for approximating the variance of a complicated estimate. J. Am. Statist. Assoc., 71, 315-321.

Yates, F. (1949) Sampling Methods for Censuses and Surveys. Griffin, London.